Nonlinear Scale Space Theory in Texture Classification Using Multiple Classifier Systems

Mehrdad J. Gangeh[1], Amir H. Shabani[2], and Mohamed S. Kamel[1]

[1] Department of Electrical and Computer Engineering, University of Waterloo,
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
{mgangeh, mkamel}@pami.uwaterloo.ca
[2] Department of System Design Engineering, University of Waterloo,
200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1
hshabani@engmail.uwaterloo.ca

Abstract. Textures have an intrinsic multiresolution property due to their varying texel size. This suggests using multiresolution techniques in texture analysis. Recently linear scale space techniques along with multiple classifier systems have been proposed as an effective approach in texture classification especially at small sample sizes. However, linear scale space blurs and dislocates conceptually meaningful structures irrespective of the type of structures exist. To address these problems, we utilize nonlinear scale space by which important geometrical structures are preserved throughout the scale space construction. This adds to the discrimination power of the classification system at higher scales. We evaluate the effectiveness of this approach for texture classification in Brodatz dataset using multiple classifier systems and learning curves. Compared with the linear scale space, we obtain higher accuracy in texture classification utilizing the nonlinear scale space.

Keywords: Multiscale, Nonlinear scale space, texture, multiple classifier systems.

1 Introduction

Texture provides important information in various fields of image analysis and computer vision. It has been used in many different problems including texture classification, texture segmentation, texture synthesis, material recognition, 3D shape reconstruction, color-texture analysis, appearance modeling, and indexing [1-4].

As texture is a complicated phenomenon, there is no definition that is agreed upon by the researchers in the field [2, 3]. This is one of the reasons that there are various texture descriptors in the literature, each of which tries to model one or several properties of texture depending on the application in hand.

However, most textures show multiresolution property. In the recent years, multiresolution techniques become prevalent in texture analysis due to this intrinsic multiscale nature of textures. Some of the most well-known multiresolution techniques on texture analysis in the literature are: multiresolution histograms [4] including locally orderless images [5], multiresolution local binary patterns [6], multiresolution Markov random fields [7], wavelets [8], Gabor filters [8, 9], multiresolution fractal feature vectors [10], texton based approaches especially those based on MR8 (maximum response 8) filter banks [11], and techniques based on scale space theory [5, 12].

Despite the success of multiresolution techniques in texture analysis, these techniques suffer from high dimensional feature space. This is due to the concatenation of the feature subsets obtained from different scales to be submitted to a classifier. This high dimensional feature space causes that the classifier suffers from the 'curse of dimensionality' [13], i.e., many data samples are required to train the classifier with a reasonable performance. This drawback is not usually revealed in the literature as the results are reported for sufficiently large training set size.

Recently, an alternative approach based on multiple classifier systems (MCS) is proposed that avoids this problem by submitting each feature subset (obtained at a resolution) to a classifier, which is called a base classifiers (BC). Hence, instead of fusion of feature subsets, the decisions made by these BCs are combined. The improvement in the results is especially significant at small sample sizes, which is shown by using learning curves [14].

Linear scale space is used in [14] as multiresolution technique. However, linear scale space suffers from two main restrictions: first, it blurs all the structures in the image without considering their geometrical meaning. This may destroy meaningful structures especially at higher scales. Second, it dislocates the structures in the image, which is due to homogeneous diffusion of the image at all directions irrespective of the structures exist. The first issue is more important in texture classification as we would like to use the information at higher scales to improve the performance of the classification system [14]; vanishing the structures at higher scales may limit this goal.

We propose using nonlinear scale space here to preserve the structures at higher scales and show that this improves the performance of the classification system in comparison to linear scale space, especially at small sample sizes.

2 Scale Space in Texture Classification

In this section, the theoretical background needed for this research is explained. Specifically linear versus nonlinear scale space theory, feature extraction, and multiple classifier systems in the context of multiresolution texture classification are discussed.

2.1 Nonlinear versus Linear Scale Space

A linear scale-space representation of an image can be derived from diffusion equation as given in (1) with constant diffusivity g and (time-like) scale variable s

$$\frac{\partial I}{\partial s} = div\ (g.\nabla I)\ . \tag{1}$$

Using convolution integral, this diffusion equation corresponds to the Gaussian smoothing of the original image I_0 with varying standard deviations. The variance of the Gaussian kernel is, therefore, proportional to the scale parameter ($\sigma^2 = 2s$). In this linear diffusion equation, the intensity of each pixel is evolved by the divergence of the radial spatial concentration gradients (∇I) of the surrounding pixels.

Any multiscale signal processing approach that uses this linear (Gaussian) scale space filtering suffers from two drawbacks. First, Gaussian smoothing is an isotropic diffusion filtering in which two (or more) regions of different structures might merge as the scale increases. In texture recognition, this side effect may result in blurring of conceptually meaningful structures such as parallel stripes shown in Fig. 1. Consequently, the extracted features at higher scales are less informative and reliable. Second, due to the dislocation of important structures such as edges, any feature extractor has dislocation problem.

To avoid undesirable blurring and dislocation of important structures (e.g., edges) in linear scale space filtering, it is proposed in [15] to control the diffusivity by incorporating the evolving image as a feedback in the smoothing process as follows:

$$\frac{\partial I}{\partial s} = div(g|\nabla I|.\nabla I) \ . \quad (2)$$

In other words, image gradient is used as a measure of edge map. Consequently, an edge-stopping function, like what is given in (3), controls the diffusivity at each direction in this anisotropic filtering scheme

$$g(|\nabla I|) = e^{-(|\nabla I|/k)^2} \ . \quad (3)$$

Stopping the diffusion at the direction of gradients higher than a threshold (k) prevents the sharing of intensity between two (or more) different regions in the image and, hence, avoids their fusion. In this way, as the scale increases, the homogeneous regions smooth more while different regions are still separated.

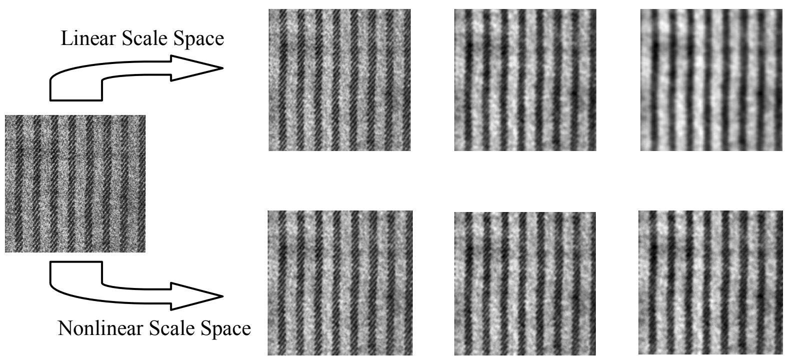

Fig. 1. Linear (*top row*) versus nonlinear (*bottom row*) scale space on texture D11 of Brodatz album (*left texture*). Note that as the scale increases (from left to right), linear scale space fails to preserve small slant patterns while nonlinear scale space can successfully do it.

2.2 Multiscale Feature Spaces

Multiscale feature subsets are obtained by computing the N-jet of derivatives up to the second order at various scales on patches. This means that in (2), I is a patch whose derivative is computed at scale s. Hence, features are computed at each scale and derivative to generate n-dimensional vectors $\mathbf{x}^{(i)} = [x_1, \dots, x_n] \in \mathbb{R}^n, i = 1, \dots, ns \times nd$, where ns and nd are the number of scales and derivatives respectively. This generates $m = ns \times nd$ feature subsets at various scales/derivatives.

These feature subsets can be composed into a single feature space $\mathbf{x} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}]^T$, which is called *distinct pattern representation* (DPR) [16].

2.3 Feature Extraction

As the pixels in each patch are used as the features, the dimensionality of the feature subsets (n) depends on the patch size. As discussed in [14], it is beneficial to the performance of the classification system to increase the patch size at higher scales. The main reason is that at higher scales the coarser structures are emphasized and hence, they should be looked at through larger windows. This increases the dimensionality of feature subsets at higher scales. There are various feature reduction techniques in the literature among which principal component analysis (PCA) is adopted in this research. It is shown in [17] that PCA can have an adaptive feature extraction effect on multiscale texture classification. That is, at higher scales where larger patches are used and, thus, the dimensionality of feature subsets is higher, PCA reduces the dimensionality more than lower scales. The reason is that due to fewer details at higher scales, fewer components are needed to preserve certain fraction of variance of the original space.

By applying PCA to original DPR, a new DPR $\mathbf{y} = [\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(m)}]^T$ is obtained in an uncorrelated space.

2.4 Multiple Classifier Systems

After computation of the DPR in reduced feature space, i.e., $\mathbf{y}$, there are two main approaches for submitting the DPR to the classification system. The common technique (see, for example, [5, 8]) is to fuse the feature subsets and submit the resulting feature space $\boldsymbol{y} \in \mathbb{R}^{k \times m}$ to one classifier[1] $D: \mathbb{R}^{k \times m} \rightarrow \Omega$, where $\Omega = \{\omega_1, \dots, \omega_c\}$ is the set of class labels for textures. The fusion of feature subsets generates a high dimensional feature space that can degrade the performance of the classifier D due to the 'curse of dimensionality' [13]. This problem is usually solved in the literature by severe dimensionality reduction of feature subsets, e.g., by

---

[1] For simplicity of the notations, here we assume that each feature subset in reduced space has a dimensionality of k and that there are m feature subsets. However, as mentioned in subsection 2.4, due to adaptive feature reduction effect of PCA, the dimensionality of feature subsets are not necessarily the same.

computation of moments of histogram [5] or estimation of energy at the output of filter banks [8].

An alternative solution is to submit the DPR to an ensemble of classifiers [14]:

$$\Gamma = \{D_1, \dots, D_m\}, \qquad \Gamma: \mathbb{R}^{k\times m} \to \Omega^m \tag{4}$$

where, $D_i: \mathbb{R}^k \to \Omega, i = 1, \dots, m,$ is the base classifier (BC) trained on each feature subset $\mathbf{y}^{(\mathrm{i})} \epsilon \mathbb{R}^{\mathrm{k}}, \mathrm{i} = 1, \dots, \mathrm{m}$. The decisions made by these BCs are subsequently fused to yield a single decision on the class of the pattern submitted for classification. Hence, the problem of finding a classifier $D: \mathbb{R}^n \to \Omega$ is converted into finding an aggregation function $\mathcal{F}$ for combining the outputs of the BCs such that $\mathcal{F}: \Omega^m \to \Omega$.

The outputs of the BCs makes a decision matrix, which is also called *decision profile* (DP) as given in (5)

$$\boldsymbol{DP}(\mathbf{y}) = \left[\begin{array}{ccccc} d_{1,1} & \cdots & d_{1,j} & \cdots & d_{1,c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{ns,1} & \cdots & d_{ns,j} & \cdots & d_{ns,c} \\ \hline d_{ns+1,1} & \cdots & d_{ns+1,j} & \cdots & d_{ns+1,c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{2ns,1} & \cdots & d_{2ns,j} & \cdots & d_{2ns,c} \\ \hline \vdots & \ddots & \vdots & \ddots & \vdots \\ \hline d_{ns\times(nd-1)+1,1} & \cdots & d_{ns\times(nd-1)+1,j} & \cdots & d_{ns\times(nd-1)+1,c} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{ns\times nd,1} & \cdots & d_{ns\times nd,j} & \cdots & d_{ns\times nd,c} \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c} d \\ \vdots \\ d \end{array}}\right\} L \\ \left.\vphantom{\begin{array}{c} d \\ \vdots \\ d \end{array}}\right\} L_x \\ \vdots \\ \left.\vphantom{\begin{array}{c} d \\ \vdots \\ d \end{array}}\right\} L_{yy} \end{array} \tag{5}$$

In (5), each row is the output of one BC and the DP is divided to some submatrices to represent the different derivatives at multiple scales. Here, we assume that the outputs of the BCs are continuous values. That is, each base classifier D_i in the ensemble generates a c-dimensional vector $[d_{i,1}, \dots, d_{i,c}] \epsilon [0, 1]^c$.

The outputs of the BCs can be combined in one stage. However, in multiscale analysis, it makes sense to group the different derivatives of the same scale in a first stage (as shown in (5) for scale S_1) and then different scales in a second stage to see the effect of each scale on the overall performance of the system. The structure of the proposed two-level classification system is shown in Fig. 2 and can be formulated as:

$$\mu_j(\mathbf{y}) = \mathcal{F}[\mathcal{G}(\mathbf{d}_1^j), \mathcal{G}(\mathbf{d}_2^j), \dots, \mathcal{G}(\mathbf{d}_{ns}^j)]\ , \tag{6}$$

where $\mathcal{G}: [0,1]^{nd} \to \mathbb{R}$ is the first aggregation function and the vector $\mathbf{d}_i^j$ is defined as:

$$\mathbf{d}_i^j = (d_{i,j}, d_{i+ns,j}, \dots, d_{i+ns\times(nd-1),j}), \qquad i = 1, \dots, ns\ . \tag{7}$$

3 Experimental Setup

To evaluate the performance of the system on the classification of small texture patches, we perform some experiments on supervised classification of several test images from Brodatz album. These test images are shown in Fig. 3. All textures used are homogeneous and have a size of 640 × 640 with 8 bit/pixel intensity resolution.

Data Preparation and Preprocessing. There is only one texture image per class in Brodatz album. Hence, to guarantee disjoint training and test sets, each image is divided into two halves. The upper and lower halves are used for the extraction of training and test patches, respectively. The patches of 32 × 32 pixels are then systematically taken from top left to bottom right of each half with some overlap to extract 1769 patches in total from each half. In order to make sure that the classification is based on the texture type not the variations in average intensity or contrast, the patches should be indiscriminable to their mean and variance. To this end, DC cancellation and variance normalization are performed on each patch.

Computation of Multiscale Patches. The N-jet of derivatives up to the second order is used for the computation of multiscale patches. For nonlinear scale space, we set experimentally the edge threshold $k = 10$ and select three scales evenly distributed in 250 iterations of nonlinear diffusion equation (2). This iteration is performed with scale difference $ds = 0.25$ and central finite difference operation. Similar to what is reported in [14], I (zero$^{\text{th}}$ order derivative), I_x, I_y, I_{xx}, I_{xy}, and I_{yy} are computed at multiple scales for each patch.

Construction of Training and Test Sets. As described in subsection 2.4 and also shown in [14], increasing the patch size at higher scales is beneficial to the performance of the classification system. Hence after preprocessing and computing the multiscale patches, the patches of 18 × 18, 24 × 24, and 30 × 30 are taken from the central part of multiscale patches at scales S_1, S_2, and S_3 respectively.

Feature Extraction. Principal component analysis (PCA) is adopted as the feature extraction technique. PCA is computed over all classes in each scale/derivative separately and 95% of original variance is retained in uncorrelated space. As discussed in 2.3, fewer components are needed to retain this percentage of variance at higher scales due to fewer details available at these scales.

Multiple Classifier System. A two-stage parallel combined classifier with the structure shown in Fig. 2 is used in the experiments. Quadratic discriminant classifier (QDC) with regularization at scale S_1 performed the best among the base classifiers (BCs) tested and hence adopted in our experiments. Regularization at scale S_1 is required because the dimensionality of feature subsets at this scale (even after using PCA) is still high and this degrades the performance of the BC at small sample sizes. The mean combiner is used at both stages as it consistently shows good performance comparing to other type of combiners over different sample sizes.

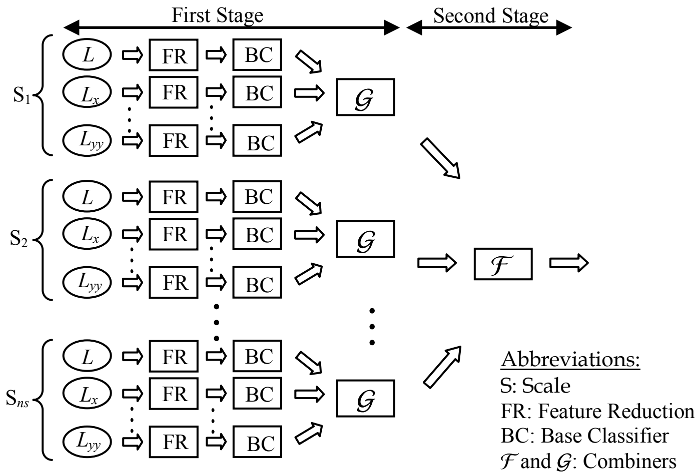

Fig. 2. The structure of proposed two-stage multiple classifier system. In first stage, different derivatives at the same scale are combined. In second stage, different scales are combined.

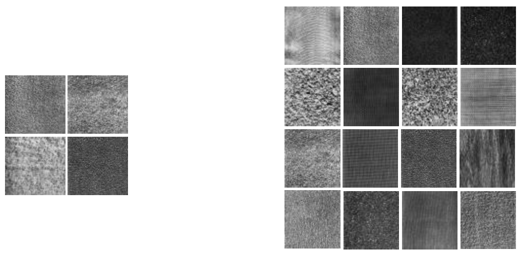

Fig. 3. 4-class (D4, D9, D19, and D57) and 16-class (D3, D4, D5, D6, D9, D21, D24, D29, D32, D33, D54, D55, D57, D68, D77, and D84) problems of Brodatz used in the experiments.

Evaluation. One of the main shortcomings of the papers in multiresolution texture classification is reporting the performance of the system at only a single (usually large) sample size. This keeps the performance of the algorithm in small training set sizes unrevealed. To overcome this problem, we use the learning curves to show the classification error of patches at various sample sizes from 10 to 1500. The experiments are repeated 5 times on different randomized patches in training and test sets and averaged results are reported. The test set size is fixed at 900.

4 Results

In this section, we present the results of texture classification using nonlinear scale space and two-stage multiple classifier systems. The performance is compared with linear scale space to show how using nonlinear scale space helps to improve the results, especially at smaller sample sizes.

The results for 4-class and 16-class problems using nonlinear and linear scale space are shown in left and right graphs of Fig. 4, respectively. The top graphs are for 4-class and bottom graphs are for 16-class problems of Brodatz dataset. In each graph, the thick curve displays the overall performance of the classification system, i.e., the output of the second stage in the proposed structure shown in Fig. 2. Thinner curves are the intermediate results, i.e., the outputs of the first stage of classification in Fig. 2, which are the results of combining different derivatives at the same scale.

Comparing the overall performance of left (nonlinear SS) and right (linear SS) graphs in Fig. 4 clearly shows the advantage of using nonlinear over linear scale space. The improvement is especially important at small sample sizes and could be due to this phenomenon that nonlinear SS preserves the structures at higher scale and this adds to the discriminative power of base classifiers at these scales. As can be seen from the graphs in Fig. 4, the performance of combined S_1 and combined S_2 are improved in nonlinear SS comparing to linear one. The overall performance is subsequently improved.

To verify the superiority of combined classifiers over combined feature space (CFS), which is the common technique in the literature, we here compare these two techniques on 4-class problem of Brodatz. Fig. 5 displays these results using CFS. Here, the feature subsets from different scales and derivatives are concatenated and after feature reduction, the combined feature space is submitted to a single QDC with the same regularization parameters as the BCs in Fig. 2. As can be seen from comparing Fig. 5 and top right graph of Fig. 4, high dimensional feature space of CFS, causes that the single classifier suffers from the 'curse of dimensionality' and its performance is significantly degraded especially at smaller sample sizes. If there are many data samples for training, we expect that CFS performs asymptotically as good as MCS. However, in this example this will happen for more than 1500 data samples which is the maximum training set size used in our experiments.

5 Discussion and Conclusion

In this paper nonlinear scale space along with multiple classifier systems are proposed for texture classification. This is to address the problem of linear scale space in blurring and dislocating the important texture structures. Consequently, we obtained improvement in classification of Brodatz texture dataset.

It is shown using learning curves and multiple classifier systems that nonlinear scale space can improve the performance of texture classification system especially at small sample sizes. This is due to more discriminative power available at higher scales of nonlinear scale space comparing to linear one. The improvement of performance at small training set size is important in applications where data acquisition is cumbersome or costly and the number of data samples for training the texture classification system is limited. This is, for example, the case in medical applications such as the diagnosis of lung diseases in high resolution CT [12] or liver diseases in B-scan images of ultrasound [10].

It is also shown that multiple classifier systems improve the performance of texture classification system based on multiresolution techniques comparing to combined feature space.

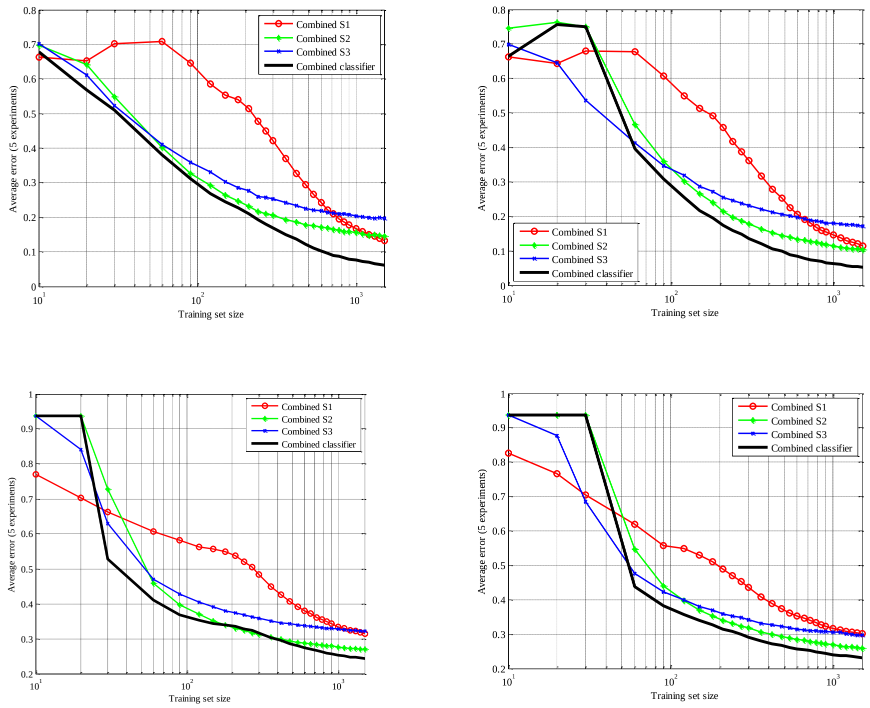

Fig. 4. Learning curves for the classification of 4-class (*top row*) and 16-class (*bottom row*) problems of Brodatz using nonlinear (*left*) and linear (right) scale space texture classification system with the structure proposed in Fig. 2.

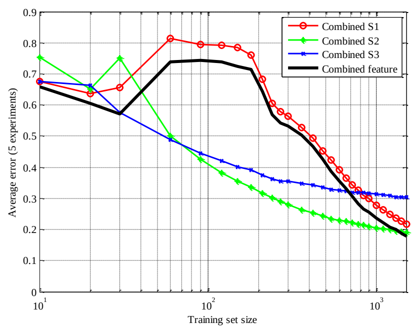

Fig. 5. The learning curves for the classification of 4-class problem of Brodatz using combined feature space (CFS) technique. These curves should be compared with the ones on top right graph of Fig. 4 which are the results for the same problem using multiple classifier systems.

Acknowledgments. The first author is funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Canada Graduate Scholarship (CGS D3-378361-2009). The first author would also like to thank Markus van Almsick from Eindhoven University of Technology, the Netherlands for useful discussions on nonlinear scale space.

References

1. Petrou, M. and Sevilla, P.G., Image Processing Dealing with Texture. John Wiley and Sons, West Sussex (2006)
2. Ahonen, T. and Pietikainen, M.: Image Description Using Joint Distribution of Filter Bank Responses. Pattern Recognition Letters, Vol. 30, 368--376 (2009)
3. _, Handbook of Texture Analysis. M. Mirmehdi, X. Xie, and J. Suri, eds., Imperial Collage Press, London (2008)
4. Hadjidemetriou, E., Grossberg, M.D., Nayar, S.K.: Multiresolution Histograms and Their Use for Recognition. IEEE Trans. on PAMI, Vol. 26, 831--847 (2004)
5. van Ginneken, B., ter Haar Romeny, B.M.: Multi-scale Texture Classification from Generalized Locally Orderless Images. Pattern Recognition, Vol. 36, 899--911 (2003)
6. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. IEEE Trans. on PAMI, Vol. 24, No. 7, 971--987 (2002)
7. Zhu, S.C., Wu, Y., and Mumford, D.: Filters, Random Fields and Maximum Entropy (FRAME): Towards a Unified Theory for Texture Modeling. International Journal of Computer Vision, Vol. 27, No. 2, 107--126 (1998)
8. Randen, T., Husoy, J.H.: Filtering for Texture Classification: A Comparative Study. IEEE Trans. on PAMI, Vol. 21, 291--310 (1999)
9. Jain, A.K., Farrokhnia, F.: Unsupervised Texture Segmentation Using Gabor Filters. Pattern Recognition, Vol. 24, 1167--1186 (1991)
10. Lee, W.L., Chen, Y.C., and Hsieh, K.S.: Unsupervised Segmentation of Ultrasonic Liver Images by Multiresolution Fractal Feature Vector. Information Sciences, Vol. 175, No. 3, 177--199 (2005)
11. Varma, M. and Zisserman, A.: A Statistical Approach to Texture Classification from Single Images. International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis, Vol. 62, 61--81 (2005)
12. Sluimer, I.C., Prokop, M., Hartmann, I., and van Ginneken, B.: Automated Classification of Hyperlucency, Fibrosis, Ground Glass, Solid, and Focal Lesions in High-Resolution CT of the Lung. Medical Physics, Vol. 33, No. 7, 2610--2620 (2006)
13. Jain, A.J., Duin, R.P.W., Mao, J.: Statistical Pattern Recognition: A Review. IEEE Trans. on PAMI, Vol. 22, No. 1, 4--37 (2000)
14. Gangeh, M.J., ter Haar Romeny, B.M., and Eswaran, C.: Scale Space Texture Classification Using Combined Classifiers. In: Ersboll, B.K. and Pedersen, K.S. (eds.) SCIA 2007, LNCS, Vol. 4522, pp. 324--333. Springer, Heidelberg (2007)
15. Perona, P. and Malik, J.: Scale space and edge detection using anisotropic diffusion. IEEE Trans. on PAMI, Vol. 12, No. 7, 629--639 (1990)
16. Kittler, J., Hatef, M., Duin, R.P.W., and Matas, J.: On Combining Classifiers. IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 20, No. 3, 226--239 (1998)
17. Gangeh, M.J., ter Haar Romeny, B.M., and Eswaran, C.: The Effect of Sub-sampling in Scale Space Texture Classification Using Combined Classifiers. In: Int'l Conf. on Intelligent and Advanced Systems (ICIAS), Kuala Lumpur, Malaysia, pp. 25--28 (2007)